HƯỚNG TỚI MÃ HÓA HÌNH ẢNH CHÍNH XÁC: CẢI THIỆN TẠO ẢNH TỰ HỒI QUY VỚI LƯỢNG TỬ HÓA VECTOR ĐỘNG

Ngô Trần Tuấn Anh - 250101003

Tóm tắt

- Lớp: CS2205.CH201
- Link Github của nhóm: https://github.com/TuananhSR/CS2205.CH201
- Link YouTube video: https://www.youtube.com/watch?v=3lse2_0KY2o
- Họ và Tên: Ngô Trần Tuấn Anh
- MSHV: 250101003

Giới thiệu

- **Vấn đề:** Mã hóa lưới cố định gây dư thừa tài nguyên ở vùng đơn giản và thiếu hụt chi tiết ở vùng phức tạp.
- **Bài toán tính toán:** Giải quyết mâu thuẫn giữa mã hóa tĩnh (fixed-length) và mật độ thông tin không đồng đều của ảnh.
- **Giải pháp trọng tâm:** Kết hợp mã hóa độ dài biến thiên (DQ-VAE) và sinh ảnh từ thô đến mịn (DQ-Transformer).
- **Input/Output:** Ảnh/Nhãn lớp →Ảnh tổng hợp sắc nét, đạt tối ưu giữa chất lượng (FID) và tốc độ (FPS).
- **Ứng dụng:** Tối ưu hóa mô hình AI quy mô lớn trên thiết bị cá nhân và hệ thống nén dữ liệu thế hệ mới.

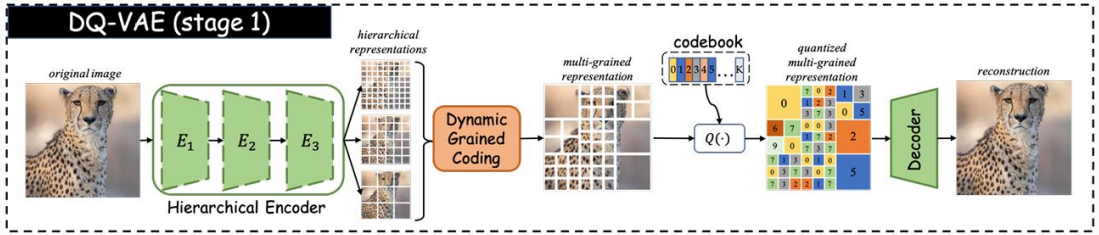
Mục tiêu

- **DQ-VAE:** Mã hóa độ dài biến thiên (DGC & Budget Loss) tối ưu theo mật độ thông tin vùng ảnh.
- **DQ-Transformer:** Sinh ảnh phân cấp (coarse-to-fine) thay thế raster-scan bằng kiến trúc Stacked Transformer.
- **Thực nghiệm:** Cải thiện ~7% chỉ số FID và tăng tốc độ suy luận trên tập dữ liệu FFHQ & ImageNet.

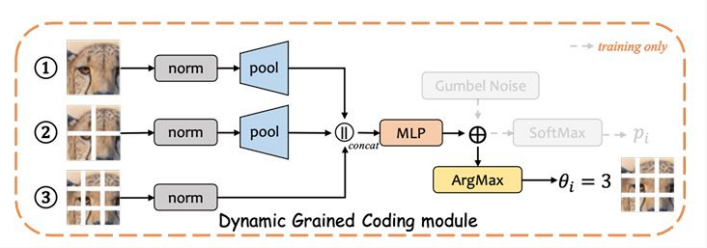
Nội dung và Phương pháp

1. Xây dựng hệ thống mã hóa hình ảnh động (DQ-VAE)

- **Mã hóa phân cấp (Hierarchical Encoder):** Trích xuất đặc trưng hình ảnh tại nhiều cấp độ chi tiết (granularities) khác nhau.
- **Mô-đun DGC (Dynamic Grained Coding):** Sử dụng mạng cổng (gating network) và kỹ thuật Gumbel-Softmax để lựa chọn cấp độ mã hóa tối ưu cho từng vùng ảnh.
- **Hàm mất mát Budget Loss:** Điều phối tỷ lệ phân bổ mã giữa các phân cấp, cân bằng giữa độ chính xác tái cấu trúc và tính nhỏ gọn của chuỗi mã.
- **Mục tiêu:** Khắc phục triệt để hiện tượng thiếu chi tiết ở vùng phức tạp và dư thừa mã ở vùng đơn giản.



Hình 1: DQ-VAE gán mã có độ dài biến thiên cho từng vùng ảnh thông qua mô-đun Mã hóa Phân hạt Động (DGC).

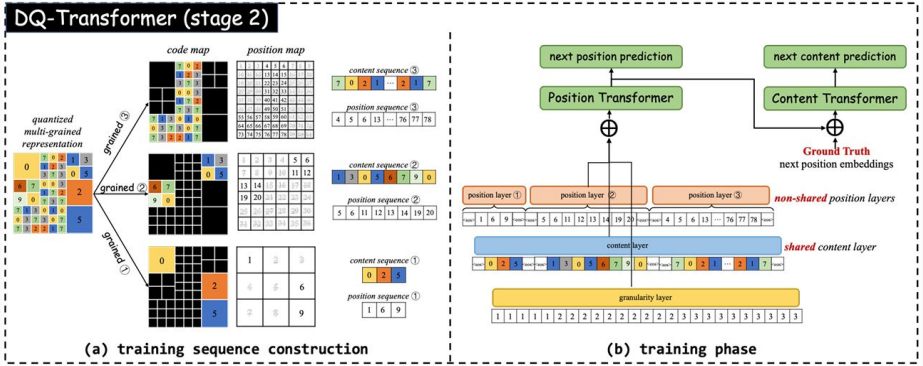


Hình 2: Minh họa mô-đun Mã hóa Phân hạt Động.

Nội dung và Phương pháp

2. Thiết kế mô hình tạo ảnh tự hồi quy phân cấp (DQ-Transformer)

- **Kiến trúc Stacked Transformer:** Dự đoán luân phiên giữa vị trí (Position-Transformer) và nội dung (Content-Transformer) của mã thông báo.
- **Thiết kế lớp đầu vào chuyên biệt:** Kết hợp lớp nhúng nội dung chung (shared-content) và lớp nhúng vị trí riêng biệt (non-shared-position) cho từng cấp độ chi tiết.
- **Thứ tự sinh ảnh Coarse-to-fine:** Ưu tiên kiến tạo khung sườn từ các vùng mịn (smooth) trước khi lấp đầy các chi tiết cục bộ (fine-grained).
- **Mục tiêu:** Thay thế quét raster-scan bằng quy trình sinh ảnh tự nhiên, đảm bảo tính nhất quán cấu trúc toàn cục.



Hình 3: DQ-Transformer mô hình hóa luân phiên vị trí và nội dung mã bằng các lớp Transformer xếp chồng, tạo ảnh tự hồi quy từ thô đến tinh.

Nội dung và Phương pháp

3. Thực nghiệm, đánh giá và so sánh hiệu năng

- **Huấn luyện quy mô lớn:** Triển khai trên tập dữ liệu FFHQ (sinh ảnh khuôn mặt) và ImageNet (sinh ảnh đa lớp theo điều kiện).
- **Đo lường định lượng:** Sử dụng các chỉ số chuẩn FID (chất lượng ảnh) và IS (độ đa dạng) để kiểm chứng hiệu quả.
- **Đánh giá hiệu suất:** So sánh tốc độ suy luận (Inference speed) với các mô hình SOTA (ViT-VQGAN, RQ-VAE) trên cùng cấu hình phần cứng RTX-3090.
- **Phân tích loại trừ (Ablation Study):** Đánh giá vai trò cụ thể của các thành phần then chốt: mô-đun DGC, Budget Loss và thứ tự sinh ảnh phân cấp.

Kết quả dự kiến

- **Hiệu quả mã hóa (DQ-VAE):** Chỉ số rFID dự kiến giảm 10-15% so với VQGAN; tái tạo chính xác hơn tại các vùng thông tin dày đặc.
- Chất lượng sinh ảnh:
  - **FFHQ:** FID đạt ngưỡng < 5.0 (cải thiện 5-8% so với ViT-VQGAN).
  - **ImageNet:** Inception Score (IS) đạt > 170, đảm bảo độ sắc nét và đa dạng.
- **Hiệu suất suy luận:** Tốc độ sinh ảnh nhanh hơn từ 1.5 – 2 lần nhờ loại bỏ mã dư thừa và tối ưu hóa lộ trình tự hồi quy.
- **Tính nhất quán:** Loại bỏ lỗi đứt gãy cấu trúc; đảm bảo sự đồng nhất giữa bố cục tổng thể và các chi tiết phức tạp (da, tóc, hoa văn).
- **Sản phẩm bàn giao:** Bộ mã nguồn PyTorch hoàn chỉnh và báo cáo phân tích chi tiết (benchmark) về tác động của mã hóa động.

Tài liệu tham khảo

- [1]. Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, Rianne van den Berg: Structured Denoising Diffusion Models in Discrete State-Spaces. NeurIPS 2021: 17981-17993
- [2]. Hangbo Bao, Li Dong, Furu Wei: BEiT: BERT Pre-Training of Image Transformers. CoRR abs/2106.08254 (2021)
- [3]. Emmanuel Bengio, Pierre-Luc Bacon, Joelle Pineau, Doina Precup: Conditional Computation in Neural Networks for Faster Models. CoRR abs/1511.06297 (2015)
- [4]. Tolga Bolukbasi, Joseph Wang, Ofer Dekel, Venkatesh Saligrama: Adaptive Neural Networks for Efficient Inference. ICML 2017: 527-536